

Weekly Report

Lu Junhua

2017 年 1 月 9 日

Done

- Finished course essay (6000 words, 18 citations).
- Learned more about the data, chat with different analysts in netease, learn about their requirements and potential requirements.
- Began a revision of tvcg paper. Prof. Wu has assign tasks for all of us.
- Start testing some algorithms on the data.

To do

- Propose and refine ideas for vis.
- Review and take an exam of advanced computer architecture.
- Revise BeXplorer.

Discussion records and potential ideas

- AI组现在在做CNN输入前的一个步骤.

变长特征如何输入模型? 变长特征仍然落在一个有限的空间里, 可以用类似霍夫曼编码的形式将他们编码成等长的编码. 以一段文字用CNN预测举例, 可以将单词从上至下排列, 这样纵向就是等长的. 然后再做滑动窗口.

他们的一个需求, 尽管个人认为实现起来有一定难度并且帮助不大. 后面会多考虑怎样打通多方面工作. 图1

- 玉泉组人工重新检查用户id, 多检索出两亿条数据. 根据之前运营提供特征进行测试, 准确率高. 但是很明确的是, 由于特征本身就是根据对应外挂数据检出, 所以准确率高是必然的. 这个特征运用到另一组直接准确率骤降.
- 了解运营组找出特征的具体应用场景, 合适的时间区间. AI组认为某些特征若有丢失直接使用-1 -2等即可.

运营找外挂, 是一种启发式的方法, 从最基本的登入升级和创建. 找出一部分, 与正常人比较, 进一步判断. 一次找完以后, 在此基础上得到后验知识(通过统计等各种手段), 作为下一次的输入. 然而

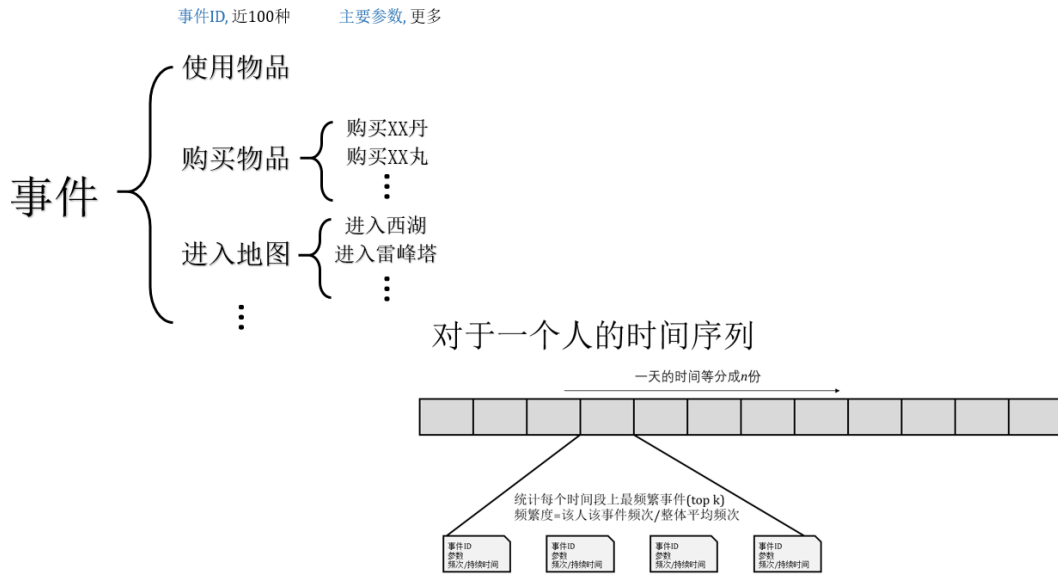


图 1: AI需求, 有考虑treemap类似方法. 如果单纯展示, 要花不少时间去做.

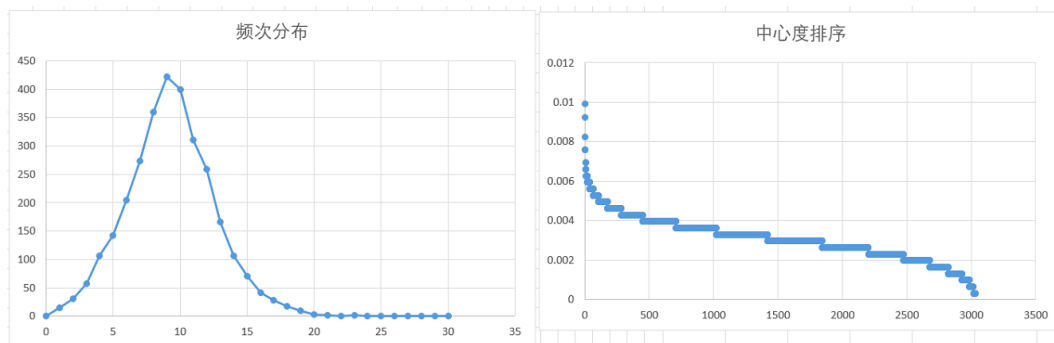


图 2: 5300075事件中当批外挂交易网络的频次与中心度分布.

这样效率是很低的, 而且就算在没有及时性的要求下, 仍然会漏掉很多外挂. 特征是会失效的, 就是说聪明的外挂它自己也一直在变化.

由于曾经见过有些文章通过金钱流动网络找挂, 但运营认为游戏中可能出现的货币流动方式过于复杂, 有200种以上, 暂时只好用收入产出比作为一种特征来算. 图2

- 与刘柏讨论的是针对TVCG review的要求, 讨论潜在case、说服力的解决方案.
- 此外, 个人负责数据问题咨询沟通事宜, 协同玉泉、我组、网易之间交流.
- 想法1: 运营部自己提出的需求, 十分类似于当初给他们的TargetVue系统, 即是选定特征, 异常打分、展示, 可视化解读、分析、改进模型. 这是最简单的一个套路. 检测算法最近也有在尝试.
- 想法2: 类似于今年VisMatchMaker, NameClarifier那样的启发式的可视化, 这两篇文章问题选的很小, 但是很精致, 完美体现人的思考与机器的结合. 如果我们能帮助原本纯手动的查找方式, 根据特征关联做一些潜在嫌疑特征推荐, 根据推荐找更多的嫌疑人. 这在一定程度上可以说是减少了人的负

担; 并且现有数据中有两批完全不同外挂名称, 两批特征. 如果开发一个系统能做到同时检出这些特征关联, 那就可以作为一个case了.

- 如果我们从另一角度看, 人肉找不可靠. 可以用某种异常检测算法检测出某些嫌疑度高的人. 这些人若包含了比人肉找的外挂以外的人, 那么论文就好说了, 可以指出一般游戏部门工作的不可靠性.
- 数据尚未处理完成, 集群也出了些小bug, 暂时先定了处理方法. 尽快处理完, 考完试会更加集中搞.